Padraig O'Brien, Data Scientist

Cork, Ireland, +353 85 1756825, Padraigobrien00@gmail.com

LINKS

Linkedin, Github

PROFILE

Data Scientist and technical lead with a proven track record of independently driving end-to-end data science projects. Adept at conceptualising, developing, and deploying production-level ML solutions—from an outlier detection model that underpinned securing our largest customer under intense time pressures to pioneering a computer vision system for quality control. Skilled in simulating and rigorously testing models, effectively communicating actionable insights to senior stakeholders, and mentoring emerging talent. Brings full-stack expertise in data collection, aggregation, analysis, and production developed through academic, personal, and industry projects, with a strong product sensibility that consistently delivers business impact.

EMPLOYMENT HISTORY

Sep 2023 — Present

Data Scientist, yieldHUB
Limerick

- Developed an AI-driven solution for detecting and measuring voiding in X-ray microchip images, leveraging classification algorithms, advanced linear algebra, computer vision techniques, and anomaly detection for high-precision analysis.
- Spearheaded a production-level outlier detection model - developing algorithms from scratch under intense time pressure - to revolutionise our platform and secure our largest client.
- Led full-stack development of an innovative computer vision processor, establishing an industry-first pass/fail system powered by a complex path-finding algorithm to enhance quality control processes.
- As project lead, developed custom statistical scripts leveraging our API to tailor solutions to customer requirements, ensuring scalable, efficient, and client-centric implementations.
- Collaborated extensively with cross-functional engineering teams to align on project goals, demonstrating strong communication and effective stakeholder engagement.
- Built a data ingestion pipeline to transform complex image and manufacturing data, further securing a major client and showcasing advanced data wrangling and pipeline development skills.
- Consulted on image processor redevelopment using Rust, enhancing system performance and optimising code efficiency.
- Provided expertise in statistical analysis by refining ANOVA applications to align with client evaluation criteria and improve decision accuracy.
- Actively mentored a data science intern and consulted company-wide on ML/AI and statistical techniques, elevating team capabilities and fostering a culture of continuous learning.
- Delivered impactful presentations on deep learning and large language models (LLMs), reinforcing a scientific mindset across teams.

Mar 2022 — Aug 2022

Data Engineering Intern, Groupon
Remote

- Designed and implemented multiple Python and SQL applications—including a data transfer interface, an automated data backup service, and a chatbot—to streamline and optimise data workflows. Successfully deployed the data transfer interface to customers, receiving positive feedback for both performance and user-friendly design.
- Gained hands-on experience with Google Cloud, Hadoop and Airflow, leveraging these Big Data tools to enhance data processing efficiency and manage large-scale, complex workflows.
- Built a strong foundation in handling extensive datasets through robust data wrangling and industry-standard practices, demonstrating advanced programming skills and delivering scalable solutions for data-intensive environments.

Mar 2020 — Feb 2022

Engineering Contractor, Cognex
Cork

- Managed and redeveloped product testing systems, overseeing component procurement, network configuration, and 19" rack setup to improve functionality and operational efficiency. Applied strategic project management to streamline testing processes, significantly enhancing system performance.
- Designed and maintained data-driven information portals for operators, centralising critical resources like quality alerts, instructions, and error procedures to reduce operator errors and increase distribution centre efficiency.
- Leveraged portals as training tools to standardise and expedite on-boarding, resulting in improved productivity and streamlined knowledge transfer across teams.

Sep 2019 — May 2023

BSc Data Science And Analytics, University College Cork

- Bachelor of Science in Data Science and Analytics from University College Cork (UCC), graduating with a 2.1 average. Developed expertise in Python, R, Java, and SQL, enabling proficiency in building machine learning solutions, data wrangling, and model deployment.
- Applied advanced machine learning techniques, including ensemble learning, hyper-parameter tuning, and feature engineering, to derive actionable insights from complex datasets.
- Academic training included dimensionality reduction (PCA) and advanced statistical methods like Bayesian Inference, Hypothesis Testing, and Monte Carlo Simulations, as well as data engineering foundations in data ingestion, ETL, and Big Data frameworks.
- Developed strong skills in cluster analysis, multivariate regression, and cross-validation techniques to ensure model robustness, equipping me to deliver reliable data-driven solutions.
- Experienced in data visualisation and statistical modelling, enabling effective data storytelling and insights communication to stakeholders.

ACADEMIC PROJECTS

Final Year Data Analytics Project

- Led an in-depth analytics project on the impact of Covid-19 on home advantage in soccer, achieving a 1.1 grade. Conducted comprehensive data scraping, extensive data preprocessing, and rigorous cross-validation to build robust datasets.
- Applied advanced statistical modelling to quantify factors influencing home advantage, analysing their significance across the pandemic timeline to uncover shifts in game dynamics and deliver actionable insights.
- Utilised multiple regression analysis to assess relationships between game factors and performance outcomes, ensuring a comprehensive, data-driven evaluation that highlighted key trends.

Multivariate Analysis

- Achieved a 1.1 grade on assignments implementing multivariate analyses, including Principal Component Analysis (PCA), linear and quadratic discriminant analysis, and hierarchical clustering, along with other advanced statistical methods.
- Processed and analysed extensive datasets, some with over 700 million data points, showcasing proficiency in handling high-dimensional data and deriving meaningful insights.

Algorithms & Data Structures

- Achieved a 100% grade by replicating Dijkstra's path-finding algorithm on a dataset exceeding 500,000 data points, successfully optimising it under stringent time constraints to reduce the run time from minutes to under one second.
- Later adapted the optimised algorithm in the workplace by incorporating zero cost nodes, delivering a creative and innovative solution to a persistent challenge in our computer vision project, displaying a clear algorithmic understanding and demonstrated my capacity to translate theoretical concepts into practical, high-impact solutions.

PERSONAL PROJECTS

Healthcare Appointment No-Show Predictor

- Developed a high-accuracy machine learning model to predict patient no-shows, achieving up to 90% accuracy through classification algorithms such as logistic regression, SVMs, random forests, and gradient boosting. Enhanced data preprocessing with advanced feature engineering and managed imbalanced datasets to ensure model robustness.
- Applied cross-validation techniques to validate reliability and optimise model performance, delivering actionable insights that improved scheduling efficiency and informed operational planning for healthcare providers. This project demonstrated the impact of data-driven decision-making on critical process improvements.

Dynamic Car Price Prediction Model

- Engineered a car price prediction model using advanced regression techniques, including Lasso, Ridge, and Gradient Boosted Regression, to deliver accurate market-based pricing assessments. Enhanced model precision by 12% through comprehensive data preprocessing, incorporating EDA, one-hot encoding, and strategic feature selection.
- Generated actionable insights that underscored the model's value in facilitating data-driven decision-making, demonstrating strong skills in data wrangling, analytics, and model refinement.